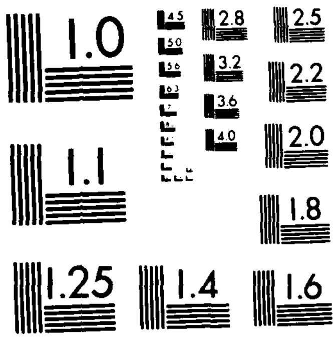
AD-A172 706 SYMPOSIUM ON PATTERN RECOGNITION METHODS IN ANALYTICAL SPECTROSCOPY (1ST) (U) UTAH UNIV SALT LAKE CITY H L MEUZELAAR AUG 86 ARO-23983.1-LS-CF DAAL03-86-G-0080 1/1

UNCLASSIFIED F/G 7/4 NL


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0

ARO 23983.1-LS-CF

(2)

AD-A172 706

FIRST SYMPOSIUM ON

PATTERN RECOGNITION METHODS IN ANALYTICAL SPECTROSCOPY

SNOWBIRD, UTAH

JUNE 16 - 18, 1986

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) ARO 23983.1-LS-CF
6a. NAME OF PERFORMING ORGANIZATION University of Utah 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION U. S. Army Research Office
6c. ADDRESS (City, State, and ZIP Code) Salt Lake City UT 84108	7b. ADDRESS (City, State, and ZIP Code) P. O. Box 12211 Research Triangle Park, NC 27709-2211
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U. S. Army Research Office 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER DAAL03-86-G-0080
8c. ADDRESS (City, State, and ZIP Code) P. O. Box 12211 Research Triangle Park, NC 27709-2211	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

First Symposium on Pattern Recognition Methods in Analytical Spectroscopy

12. PERSONAL AUTHOR(S)

Henk L. C. Meuzelaar

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final Report	FROM 1 Apr 86 TO 31 Mar 87	August 1986	

16. SUPPLEMENTARY NOTATION

The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The goal of the meeting was to provide a direct interface between the fields of Analytical Spectroscopy (e.g., MS, IR, NMR, UV-VIS) and Chemometrics. The following topics fall within the scope of the program: Spectral Enhancement and Deconvolution, Data Reduction and Compression, Library Storage and Retrieval, Cluster and Classification Analysis, Factor and Discriminant Analysis, Multicomponent ("mixture") Analysis, Automated Spectral Interpretation, Multisource Data Base Integration, and Modeling and Prediction. Keynote lectures will be published in the form of a hardcover book by Plenum Publishing Company.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code) / 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

First Symposium on
PATTERN RECOGNITION METHODS IN ANALYTICAL SPECTROSCOPY

JUNE 16-18, 1986

SNOWBIRD, UTAH

QUALITY INSPECTED 4

Accesion For	
NTIS CRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

International Committee

Thomas L. Isenhour, Chairman
Jean T. Clerc, Switzerland
Hal J.H. MacFie, United Kingdom
Edmund R. Malinowski, USA
Charles L. Wilkins, USA

Organizing Committee

Henk L.C. Meuzelaar, Chairman
Willem Windig, Treasurer
Melinda Van, Secretary

Symposium sponsored by the Biomaterials Profiling Center, University of Utah and the U.S. Army Research Office.

OBJECTIVES AND SCOPE

The goal of the meeting is to provide a direct interface between the fields of Analytical Spectroscopy (e.g., MS, IR, NMR, UV-VIS) and Chemometrics. The following topics fall within the scope of the program:

- Spectral Enhancement and Deconvolution
- Data Reduction and Compression
- Library Storage and Retrieval
- Cluster and Classification Analysis
- Factor and Discriminant Analysis
- Multicomponent ("mixture") Analysis
- Automated Spectral Interpretation
- Multisource Data Base Integration
- Modeling and Prediction

FORMAT

The three day schedule is patterned after the well-known Gordon Research Conference format with morning and evening lecture sessions and afternoons available for social interactions or recreational activities. Each morning or evening session features only two keynote speakers with ample time for discussions. Keynote lectures will be published in the form of a hardcover book by Plenum Publishing Company. A copy of this book will be sent to register participants. Authors of papers presented in poster form are encouraged to submit full manuscripts to Analytica Chimica Acta.

GENERAL INFORMATION

All meetings and meals will be held in the Snowbird Center as specified in the following agenda. Snowbird's facilities are available to all registered participants. Snowbird is the perfect environment for relaxed, inspired meetings. It is a self-sufficient mountain hideaway, with shops, restaurants and lodging, all within walking distance of each other. Each lodge features outdoor heated swimming pools and saunas. We will take the Snowbird Aerial Tram up Hidden Peak (11,000 ft) on Tuesday for lunch.

For those who registered for the spouse program, an additional handout will be provided explaining where you will go and what you will do. The spouse program is scheduled for Monday and Wednesday from 9:00 a.m. to 2:00 p.m.

Meals are included in the accomodation fee. At the time of registration each participant and paying guest will receive a name badge with a sticker to indicate payment of the accomodation fee. Please have this badge available at mealtime. If you are not purchasing the accomodation package, you may purchase tickets for individual meals. These tickets can be purchased at the registration desk.

If you have any questions please feel free to call upon any of the representatives of the Biomaterials Profiling Center.

First Symposium on PATTERN RECOGNITION METHODS IN ANALYTICAL SPECTROSCOPY

AGENDA

DATE/TIME	FUNCTION	LOCATION*
<u>Sunday, June 15, 1986</u>		
	Registration**	Deck
7:00 - 9:00 p.m.	"Wine and Cheese" Reception for Registered Symposium Participants and Spouses	Deck
<u>Monday, June 16, 1986</u>		
7:30 - 8:00	Registration	Plaza
8:00 - 8:45 a.m.	Breakfast Buffet	Alpine South
8:45 - 9:00 a.m.	<u>Welcome and Update - Henk L.C. Meuzelaar</u>	Plaza
<u>Morning Session - Thomas L. Isenhour, Chairman</u>		
9:00 - 10:00 a.m.	<u>Keynote Speaker - Peter R. Griffiths</u> "Spectral Enhancement and Deconvolution Techniques in Infrared Spectroscopy"	Plaza
10:00 - 10:15 a.m.	Coffee Break	Deck
10:15 - 11:15 a.m.	<u>Keynote Speaker - Stephen R. Heller</u> "Library Storage and Retrieval Methods in Infrared Spectroscopy"	Plaza
11:15 - 12:00 a.m.	Posters & Demonstrations	Alpine North
12:00 - 1:00 p.m.	Lunch Buffet	Alpine South
6:00 - 7:00 p.m.	Dinner Buffet	Alpine South
<u>Evening Session - Paul C. Painter, Chairman</u>		
7:30 - 8:30 p.m.	<u>Keynote Speaker - Hugh B. Woodruff</u> "Novel Applications of Pattern Recognition and Knowledge-Based Methods in Infrared Spectroscopy"	Plaza
8:30 - 8:45 p.m.	Coffee Break	
8:45 - 9:45 p.m.	<u>Keynote Speaker - Abraham Savitzky</u> "Applications of Pattern Recognition Methods in Infrared Spectroscopy"	Plaza

* All meeting rooms are located on the second floor of the Snowbird Conference Center (see enclosed map). The main entrance to the meeting rooms is "Plaza Restaurant".

** The registration desk will be manned throughout the symposium 8:00 a.m. to 12:30 p.m. and from 5:30 p.m. to 7:00 p.m.

Continued

DATE/TIME	FUNCTION	LOCATION
Tuesday, June 17, 1986		
8:00 - 9:00 a.m.	Breakfast Buffet	Alpine South
Morning Session - Piet J. Kistemaker, Chairman		
9:00 - 10:00 a.m.	Keynote Speaker - Edmund R. Malinowski "Evolutionary Factor Analysis in Analytical Spectroscopy"	Plaza
10:00 - 10:15 a.m.	Coffee Break	Deck
10:15 - 11:15 a.m.	Keynote Speaker - Willem Windig "Numerical Extraction of Chemical Components from Pyrolysis Mass Spectra by Multivariate Techniques"	Plaza
11:15 - 12:00 a.m.	Posters & Demonstrations	Alpine North
12:00 - 1:00 p.m.	Lunch (Box Lunch at the Top of Hidden Peak)	Tram Room
6:00 - 7:00 p.m.	Dinner - Western Barbecue	Deck
Evening Session - Willem Windig, Chairman		
7:30 - 8:30 p.m.	Keynote Speaker - Hal J.H. MacFie "Novel Applications of Pattern Recognition and Knowledge-Based Methods in Mass Spectrometry"	Plaza
8:30 - 8:45 p.m.	Coffee Break	
8:45 - 9:45 p.m.	Keynote Speaker - Carla Wong "Development of an AI-Based Autotuning System for Tandem Mass Spectrometry"	Plaza

Continued

DATE/TIME	FUNCTION	LOCATION
Wednesday, June 18, 1986		
8:00 - 9:00 a.m.	Breakfast Buffet	Alpine South
Morning Session - Steve Morgan, Chairman		
9:00 - 10:00 a.m.	Keynote Speaker - Tormod Naes "Multivariate Calibration: Quantification of Harmonies and Disharmonies in Analytical Data"	Plaza
10:00 - 10:15 a.m.	Coffee Break	Deck
10:15 - 11:15 a.m.	Keynote Speaker - Jean T. Clerc "Automated Spectral Interpretation Methods"	Plaza
11:15 - 12:00 a.m.	Posters & Demonstrations	Alpine North
12:00 - 1:00 p.m.	Lunch	Alpine South
6:00 - 7:00 p.m.	Dinner (Western Barbecue)	Deck
Evening Session - Henk L.C. Meuzelaar, Chairman		
7:30 - 8:30 p.m.	Keynote Speaker - Hal J.H. MacFie "Carbon-13 Nuclear Magnetic Resonance Spectrum Simulation"	Plaza
8:30 - 8:45 p.m.	Coffee Break	
8:45 - 9:45 p.m.	Keynote Speaker - Charles L. Wilkins "Synergistic Use of Spectral Data for Structural Elucidation"	Plaza
Thursday, June 19, 1986		
7:00 - 8:30 a.m.	Breakfast*	Alpine South
	CHECKOUT AND RETURN TO AIRPORT	

* Early morning travelers should arrange for special transportation to SLC at the registration desk.

Lecture Abstracts

SPECTRAL ENHANCEMENT AND DECONVOLUTION TECHNIQUES IN INFRARED SPECTROSCOPY

Peter R. Griffiths
Department of Chemistry
University of California
Riverside, California 92521

Infrared spectra are often composed of a complex blending of several broad absorption bands. When these bands are separated by more than their full width at half height, many conventional techniques for multicomponent analysis may be applied to determine the areas of each component band. These include curve fitting and several of the matrix methods, such as the K-and P-matrix. As band overlap becomes greater, solutions using these algorithms become less unique. Indeed we have found that under certain circumstances, least-squares curve fitting routines cannot even be used to fit a synthetic spectrum composed of several overlapping Lorentzian bands.

Under these circumstances it is necessary to operate in the spectrum in some way to force it to become more unique. Calculation of an even-order deviation spectrum is one way of reducing the widths of bands, but without a prior knowledge of the number of bands contributing to a given multiplet, derivative spectra can often be quite difficult to interpret a priori because of the presence of secondary lobes. We have found that deconvolution of band shapes in the Fourier domain is a very powerful way of reducing bandwidths without introducing side-lobes. Indeed after deconvolving a spectrum, curve-fitting routines can be applied to give accurate estimates of the areas of all component bands even when no calibration data are available.

Examples of the use of this technique will be given both using synthetic mixtures of nitriles (to validate the methodology) and the spectra of coals and coal extracts, where the number of bands and their true wavenumbers are unknown.

LIBRARY STORAGE AND RETRIEVAL METHODS IN INFRARED SPECTROSCOPY

Stephen R. Heller and Stephen R. Lowry
USVA-ARS, ASRI-MDCL
Room 233-C
Bldg. 005 BARC-West
Beltsville, Maryland 20705

Infrared spectroscopy (IR) is the oldest and most used spectral analysis method employed by organic chemists. This well-known technique has evolved from prism instruments to grating instruments and most recently to Fourier Transform (FT) instrumentation. As a result of both its long use and different instrumental techniques, a number of databases and search techniques have been developed over the past few decades. This presentation will describe the valuable and unique features of IR, the methods of IR spectral storage and IR data handling, along with the methods developed for searching IR databases. Included will also be a discussion of the issues involved with quality assurance and quality control of IR spectral data, and how IR spectral data, combined with chemical structure searching and other physical properties, can combine to provide the chemist with a very powerful laboratory analysis and identification system.

NOVEL APPLICATIONS OF PATTERN RECOGNITION AND KNOWLEDGE-BASED METHODS IN INFRARED SPECTROSCOPY

Hugh B. Woodruff
Merck Sharp & Dohme Research Laboratories
Division of Merck & Co., Inc.
P.O. Box 2000, Rahway, New Jersey 07065

While not the only tool used by structure elucidation chemists, infrared (IR) spectroscopy does provide a unique fingerprint of compounds and has proven to be extremely valuable. Most chemists are not sufficiently expert in the field of IR spectroscopy to be able to interpret spectra without the use of an aid such as a correlation chart. In recent years, the computer has developed into a powerful tool to aid the chemist in interpreting IR spectra.

The increased power of computers and easier and cheaper availability of mass storage on these computers have enabled scientists to take advantage of sophisticated spectral library searching algorithms. Because an IR spectrum provides a unique fingerprint of a compound, an exact match during a library search results in a virtually positive identification of the unknown. However, even the most complete spectral library contains spectra of only a small percentage of the millions of known compounds. In addition, many infrared samples are not pure compounds, hence a search may prove helpful, but certainly does not assure identification of the unknown.

For these reasons, computerized IR interpretation techniques are an important area of chemical research. A diversity of pattern recognition techniques have been investigated to aid the scientist interpret IR spectra. An overview of these approaches will be presented in this paper.

A somewhat different approach to computer-assisted IR interpretation is through the use of knowledge-based systems. Although these programs are frequently called expert systems, perhaps a better name would be a smart assistant. One such program, PAIRS, was introduced in 1980. In subsequent years, PAIRS has been used successfully by a large number of scientists. PAIRS consists of an interpreter and an extensive collection of interpretation rules. In recent years, a number of enhancements have been incorporated into PAIRS. These enhancements will be detailed in this paper.

APPLICATIONS OF PATTERN RECOGNITION METHODS IN INFRARED SPECTROSCOPY

Abraham Savitzky
Silvermine Resources, Inc.
Wilton, Connecticut 06897

An important facet of current spectral identification and search systems is that they are intended for use by persons who are not expert spectroscopists. An expert narrows the scope of his search by first identifying patterns in the spectrum, the functional groups, then focuses attention on these compounds when searching the library for matching spectra. The search for the structural units is similar, in many respects, to a search for the components of a mixture. The effect of the prefilter is to eliminate materials which accidently match the peaks of the unknown but are chemically unrelated. A significant factor in this search mode is the guidance the searcher receives when the spectrum is not contained in the search library. A *recognition system* of this type is described as one of the earliest examples of a commercially successful expert system.

The personal computer explosion has been fueled not only by the low cost of the hardware, but by the low cost and increasing sophistication of the software packages that are available. Since our search libraries constitute a unique database, it is worth asking whether standard database packages, as well as expert system development packages, can make a contribution. Some interesting results of this investigation are presented.

EVOLUTIONARY FACTOR ANALYSIS IN ANALYTICAL SPECTROSCOPY

Edmund R. Malinowski
Department of Chemistry and Chemical Engineering
Stevens Institute of Technology
Castle Point
Hoboken, New Jersey 07030

Factor analysis is a computational tool for solving multidimensional problems in analytical spectroscopy. It can be used to analyze unknown mixtures of an unknown number of unknown components. Abstract factor analysis (AFA) reveals the number of spectroscopically visible components. Target factor analysis (TFA) verifies the presence or absence of suspected components. Evolutionary factor analysis (EFA) takes advantage of experimental variables that control the evolution of components, revealing not only the concentrations of the components but also their spectra, even when there are no unique concentrations or spectral regions. The method is applied to model studies involving circular-dichroism spectra, spectra exhibiting both positive and negative intensities.

NUMERICAL EXTRACTION OF CHEMICAL COMPONENTS FROM PYROLYSIS MASS SPECTRA BY MULTIVARIATE TECHNIQUES

W. Windig and H.L.C. Meuzelaar
Biomaterials Profiling Center
University of Utah
391 S. Chipeta Way
Salt Lake City, Utah 84108

In the last 5 years pyrolysis mass spectrometry (Py-MS) has evolved from a fingerprinting method to a sophisticated method for chemical analysis of complex organic materials. Due to the complexity of Py-MS data and the lack of reference spectra, special data analysis methods had to be developed in order to retrieve the chemical information. These methods vary from a purely graphical interactive method to rotate the results of factor and discriminant analysis to a mathematical procedure to extract the spectra and the absolute concentrations of the pure components from data sets of complex mixtures without using calibration data.

In order to provide a user-friendly interface between computer and analytical chemist, these novel methods are primarily based on geometrical representations of the output of factor and discriminant analysis rather than on mathematical expressions. The purpose of the lecture is to highlight the capabilities of multivariate analysis techniques such as factor and discriminant analysis for retrieving chemical information from spectra of complex mixtures.

Applications of these techniques on MS data from biopolymer mixtures, grass leaves, lignites and jet fuels will be discussed. Some of these applications involve the analysis of a set of samples, others the analysis of a single sample using time-resolved MS data.

NOVEL APPLICATIONS OF PATTERN RECOGNITION METHODS IN MASS SPECTROMETRY

Dr. Hal Macfie
Food Research Institute
Landford, Bristol BS18 7DY
United Kingdom

The multivariate methods of three mode principal components and genralized procrustes analysis are discussed.

The use of principal components analysis and related techniques to analyse two-dimensional arrays of MS data is now widespread. However many data arrays are now three way. For example we may characterise a time profile for each mass of each spectra. The matrix formulation and geometry of 3 mode principal components analysis will be discussed and illustrated using published GC-MS data.

Generalized procrustes analysis is a technique to compare and average measurements from different sources or instruments. The matrix formulation, geometry and interpretation of the output of this technique will be discussed.

The method is used to compare classifications of organisms using conventional and chemical tests.

DEVELOPMENT OF AN AI-BASED AUTOTUNING SYSTEM
FOR TANDEM MASS SPECTROMETRY

Carla M. Wong, Hal R. Brand, and Hugh R. Gregg
Lawrence Livermore National Laboratory
P.O. Box 808
L-365
Livermore, California 94550

Triple quadrupole mass spectrometers (TQMS) are very complex, computer-controlled, multiparametric instruments which require selective tuning or optimization of over 30 operational parameters in five different operational modes. They generate incredible amounts of multidimensional data and require considerable expertise for both operating the instrument and interpreting the data. This expertise is the kind of knowledge that can be represented as procedures, or rules, of the type described in artificial intelligence (AI) research. In this environment, it is possible to encode a tuning procedure, including heuristics, to describe real-time optimization of the data acquisition process throughout the entire mass range of the TQMS. Now the tuning of a mass spectrometer no longer has to be limited to the traditional "average tuning" where the sensitivity in both the high and the low mass ranges is compromised to achieve the instrument "tuned state". Another advantage to this approach is that the instrument can be controlled in a matter such that only that data most relevant to the experiment will be collected. This ability to optimize instrument operational and data acquisition parameters while actually running an experiment has a number of advantages; the most important of which is the ability to redefine the data you want to collect next. This means that experiments become information driven rather than just data driven. This TQMSTUNE expert system enables us to optimize the instrument operational parameters based on rules associated with peak shape, intensity, resolution, interactions between tuning parameters and compound and mass differences. The rules actually extend the normal method of tuning the TQMS in MS/MS operational mode, simply because it is too time consuming to achieve this level of optimization manually. An expert system can do this unattended and have the optimized files ready to access in real time as a sample is being analyzed.

MULTIVARIATE CALIBRATION: QUANTIFICATION OF HARMONIES AND DISHARMONIES IN ANALYTICAL DATA

Tormod Naes and Harald Martens
Norwegian Food Research Institute
P.O. Box 50
N-1432 Aas-NLH
Norway

Determination of chemical concentrations can be made more rapid and reliable by combining information from several measurement variables, e.g. light absorbances at several wavelengths. Systematic "errors" can thereby be eliminated so that very unspecific measurement data can be used for quantitative determinations. Automatic warning of unexpected errors is also possible. Different classes of multivariate calibration methods will be discussed as well as different error detection methods. The methods will be illustrated by examples.

AUTOMATED SPECTRAL INTERPRETATION METHODS

Jean T. Clerc
University of Berne
CH-3012 Berne, Switzerland

All automated spectra interpretation systems hither to described in the scientific literature use the same basic algorithm, which consists of the following 5 steps:

1. Inference of partial structures from selected spectral features.
2. Construction of consistent sets of structural features.
3. Assembly of meaningful chemical structures.
4. Spectra prediction.
5. Spectra comparison.

The systems differ widely in the way these steps are implemented and in relative weight assigned to each step. To illustrate the possiblities and limitations, some selected systems will be discussed within the framework of the basic algorithm.

Library search systems represent a special case of automated interpretation systems, where the first two steps of the basic algorithm are skipped. Instead, all compounds in the library are considered as candidates, and the stored reference spectra make spectra prediction trivial. This leaves spectra comparison as the critical step. The system may either assume, that the reference library includes a compound identical to the unknown at hand (identity search), or it may focus the attention on reference compounds structurally similar to the unknown (similarity search). Which of these two strategies dominates is determined by the similarity measure used and by the type of spectral features it is based upon. The two strategies also call for differently structured libraries. Objective evaluation of the performance of automated spectra interpetation systems is today very difficult, if not impossible at all. Attempts to measure and compare the performance of library search systems for infrared spectroscopy will be discussed and preliminary results will be presented.

CARBON-13 NUCLEAR MAGNETIC RESONANCE SPECTRUM SIMULATION

Peter C. Jurs, Debra S. Egolf
Department of Chemistry
152 Davey Laboratory
The Pennsylvania State University
University Park, Pennsylvania 16802

Carbon-13 nuclear magnetic resonance specctroscopy is a powerful tool for organic structure elucidation because the signals are directly related to the surroundings of the skeletal carbon atoms. Modern NMR spectrometers generate very large quantities of data rapidly, which has increased the demand for tools to aid the spectroscopist in data analysis.

Spectral simulation techniques comprise one category of analysis methods that have proven useful in structure elucidation studies. These methods can be used to simulate the C-13 NMR chemical shifts for each candidate structure being considered as a possible solution to the structural problem. The most widely used simulation technique involves the construction of linear models relating chemical shifts to structural parameters.

We have implemented and used an interactive software system that enables the chemist to develop and apply linear models for C-13 NMR chemical shifts, using computer-generated structural parameters. The system supports the entry and storage of chemical structures and associated spectra, calculation of a variety of structural parameters, calculation and storage of linear models, and prediction of the shifts of unknown compounds. The system can compute the topological environment of carbon centers, which allows the automated selection of carbon centers for inclusion in the model formation step of a study. When predicting the shift for a carbon center in an unknown compound, the system has the capability to choose which of many stored linear models is most suitable for the prediction.

The results from our latest studies using the system will be described. One study deals with a set of 32 alkyl- and hydroxy-substituded cyclopentanes. A related study involves a data set with less structural diversity, namely 15 cyclopentanes carrying one through five hydroxyl groups. Linear models for predicting chemical shifts from structural descriptors have been developed based on carbon atom subgroupings by connectivity. The 32 compounds contain 35 unique primary carbon centers, 82 unique secondary carbon centers, 47 unique tertiary centers (36 with alkyl substituents and 11 with attached hydroxy groups), and 13 unique quaternary carbon centers. Models have been constructed for these groups based on computed structrual descriptors. Comparisons of the predicted chemical shifts and the actual observed values will be given.

SYNERGISTIC USE OF SPECTRAL DATA FOR STRUCTURAL ELUCIDATION

David A. Laude, Jr. and Charles L. Wilkins
Department of Chemistry
University of California, Riverside

Currently available computer-readable spectral databases include mass spectrometric, infrared, and nuclear magnetic resonance libraries. A number of approaches to the use of such libraries, including library search, pattern recognition, and spectral simulation (followed by library comparisons) have been developed. In recent years significant efforts have been directed toward development of algorithms and analytical systems capable of exploiting the complementary nature of these types of spectrometry. The current state of both databases and algorithms will be discussed.

Of particular recent interest is the use of quantitative and edited nuclear magnetic resonance data for synergistic interpetation of GC/IR and GC/MS library search results. This is particularily so for the identification of unknowns. The present status of research in this area will be discussed within that context.

Poster Abstracts

COMPUTER ASSISTED INTERPRETATION OF PYROLYSIS MASS SPECTRA OF TWO OIL SHALES AND THEIR CORRESPONDING KEROGENS

T. Chakravarty, W. Windig, K. Taghizadeh and H.L.C. Meuzelaar
Biomaterials Profiling Center
L.J. Shadle
Morgantown Energy Technology Center

Green River (Colorado) and Devonian (Albany) oil shale samples as well as their corresponding kerogen isolates, were analyzed using Curie-point pyrolysis mass spectrometry (Py-MS) coupled with multivariate data analysis using the SIGMA program. A time-integrated Py-MS mode was used to bring out the similarities and the differences between the shales and kerogens at different final temperatures of the Curie-point filaments. Pyrolysis mass spectra of whole oil shales and their corresponding kerogens were obtained in triplicate at six different temperatures, 358, 480, 510, 610, 770 and 980°C respectively. Factor and discriminant analysis of the resulting data were performed in order to reduce the apparent dimensionaltiy and to help reveal underlying structural details as well as differences between samples.

Numerically extracted spectra of the Colorado and Albany oil shales and kerogens at 610°C revealed characteristic differences in the composition of the pyrolyzates. For example, the three discriminant functions representing 98% of the variance of the data obtained at 610°C, showed that major differences (90% of total variance) exist between the Colorado and Albany oil shales whereas, the oil shales and their corresponding kerogens are nearly indistinguishable (differences amounting to 3 to 5% of total variance).

Pyrolyzates of the Albany samples were found to be marked by higher sulfur compounds and aromatic hydrocarbons such as benzenes, indenes and naphthalenes. However, the pyrolyzates of the Colorado samples appeared to be richer in aliphatic and alicyclic hydrocarbons. Whereas the differences in sulfur content may be assumed to reflect differences between the original depositional environments, the observed differences in aromaticity could well be due to different degrees of maturation.

Compared to the whole shale samples, the kerogens exhibited increased HCL^+ signals (apparently derived from the demineralization solvent), increased elemental sulfur and SO_2^+ signals (probably derived from the former mineral matrix) and decreased intensities of some small molecules which may have been extracted by the demineralization procedure.

Poster No. 2

THE ANALYSIS OF MICROBIAL DATA FROM PYROLYSIS MASS SPECTROMETRY IN THE CLINICAL LABORATORY

Robert Kajioka
Public Health Labs
P. O. Box 9000 Terminal A
Toronto, Canada

Automated microbial identification scans a massive library of clinically important microorganisms. Since pyrolysis and data acquisition occupy less than one-half minute, computerized data analysis could present a rate limiting step in routinely processing a large number of samples.

Results using supervised and unsupervised learning techniques, feature weighting, data selection etc. are presented. They illustrate some aspects encountered with clinical microbial isolates.

For identification of an unknown isolate a minimum number of features may act as markers to dissect out the relevant portion(s) of the library database. It is immaterial that groups belonging to the same family appear unrelated or contain unrelated strains. For specific matching of isolates against library strains, identification should, for practical purposes, coincide with the classification established by traditional methods such as serology. Since no two isolates are likely to be exactly alike except in an epidemic, difficulties may arise. Inherent pyrolysis data structure may tend to promote a unique pyrogram classification. Although the traditional method may yield a less logical classification it likely provides clinically relevent information owing to its wider use. This implies that pattern recognition be directed away from inherent overall data structure to feature selections that force data vectors into categories coincident with sero-grouping or other classical approaches. This may not be possible in some cases. The ideal solution would ultimately be a tradition of classification based on pyrolysis fingerprints.

In epidemiology strain relatedness is more significant than precise identification. Pattern recognition needs to be honed for fine distinctions. This may create problems such as false groupings based on day-to-day variations.

MULTIVARIATE ANALYSIS OF MS/MS SPECTRA OF MIXTURES OF "ISOBARIC" IONS

S. Kornig, R. Hoogerbrugge and P. G. Kistemaker
FOM Insititute for Atomic and Molecular Physics
Amsterdam, The Netherlands

In mass spectra of complex mixtures most mass peaks are not "pure". This means that various ions with the same nominal mass, but with different structures contribute to one mass peak.

In this presentation we explore the potentials of collision induced dissociation (CID) mass spectrometry in combination with multivariate data analysis to retrieve the different ion structures present in one mass peak. To facilitate this approach one has to record a number of CID spectra for mixtures with different concentrations of the ion structures present.

Changes in the concentrations can be obtained by various methods. We have exploited two approaches. In the first case CID spectra were recorded at different temperatures of the sample. In the second example we used the fact that the ion concentrations are different at the left side, the central part and at the right side of a mass peak recorded at nominal resolution.

A principal component analysis of these data-sets followed by a dedicated rotation of the principal components generates the CID spectra of the individual ion structures.

A sulphur-rich coal sample was heated in 10 s to a temperature of 600° C. During this heating period CID spectra of mass 60 ions were recorded at intervals of 2 s. From the set of 5 spectra, it was found that carbonyl sulphide, acetic acid and C5 were present in the peak at m/z 60. From a mixture of acetaldoxime and diethylether, ions were generated at a nominal mass 59.

The mass difference between the ions from the two components is 0.012 AMU (requiring a resolution of 5000 for separation). By taking a series of CID spectra from different parts of the m/z 59 peak, we could identify two ion structures. This was obtained with a spectrometer with a resolution of only 500.

SIMCA PATTERN RECOGNITION OF MASS SPECTRA OF TOXIC ORGANIC COMPOUNDS

Donald R. Scott
U. S. Environmental Protection Agency
Environmental Monitoring Systems Laboratory
Research Triangle Park, North Carolina 27711

The Shannon information content of the binary encoded mass spectra of 78 toxic organic compounds have been calculated. The information content of the full intensity mass spectral data also has been computed. The 17 masses with highest information content were used for pattern recognition studies. The pattern recognition study of the mass spectra of the 78 compounds resulted in the determination of four classes of compounds. These included aromatics without chlorine substitution, chloroaromatics, bromoalkanes and alkenes, and chloroalkanes and alkenes. Alkenes and alkanes with both chloro- and bromo-substitution were classified as bromo-compounds. The principal component models generally consisted of only one component per class, with five masses per class. However, the total alkene and alkane class had two components with twelve masses. Classification accuracy was 96% for the total aromatics and total alkanes and alkenes and 82% for the four subclasses. The importance of the binary encoded mass spectra in the successful application of SIMCA pattern recognition studies of the 78 compounds will be discussed. The relationship between the binary encoded mass spectra and a geometrical representation in a multidimensional Hamming space also will be discussed.

This is an abstract of a proposed presentation and does not necessarily reflect EPA policy.

PYROLYSIS/MASS SPECTROMETRY APPLIED TO AGRICULTURAL MATERIAL

Grant Gill Smith, B. Austin Haws, William F. Campbell,
Kay H. Asay, and John Evans
Utah State University
Logan, Utah 84322

J. J. Boon
FOM Institute for Atomic and Molecular Physics
Amsterdam, The Netherlands

Pyrolysis/mass spectrometry has been used to distinguish biotypes in three agricultural materials: insects, weeds and grasses.

Entomologists have recently recognized apple maggots (Rhagoletis pomonella) exhibit unusual behavior and have found apple maggot larvae developing in Utah cherries but not Utah apples. They have attributed this to biotypes. Entomologists can not readily distinguish the adult apple maggot biotypes through taxonomy. Pyrolysis/mass spectometry showed that the apple maggot which hosts on Utah cherries was different from the apple maggot from three other states hosting on apples. The Western cherry fruit fly (Rhagoletis indifferens) was shown to be different than the apple maggot also hosting on Utah cherries. Pupae were freeze dried soon after collection for Py/MS studies.

The United States has been infested with a rapidly spreading weed known as leafy spurge (Euphorbia esula) from Europe. There are many biotypes of this weed that are *indistinguishable with careful* morphological studies. Insects, however, can readily distinguish the biotypes by examining the chemicals in the plant, e.g., the milk or latex found in the stem. A preliminary study by Py/MS on the latex has also been able to distinguish the biotypes grown in the same greenhourse grown from roots taken from Hungary, Canada and two sites in the United States.

Py/MS studies have been used to distinguish range grasses from one another and from their hybrids and grasses which are resistance from those which are susceptible to the grass bugs, (Labops hesperius and Irbisia brachycera). This study is directed to improving the range grasses in the Western United States.

FACTOR-DISCRIMINANT ANALYSIS

Detection of Chemical Components Using the BIPLOT Technique

A.C. Tas, J. de Waart, J. van der Greef,
J. Bouwman and M.C. ten Noever de Brauw
TNO-CIVO Food Analysis Institute
Zeist, The Netherlands

Principal component analysis enables the factorization of a data matrix X into a factor scores matrix F and a factor loadings matrix A. The F-matrix contains the object coordinates on the new axis; the A-matrix represents the correlation coefficients between the new and the original axis when both X and F are standardized. This factorization opens the possibility of displaying on the plane of two principal components, objects as well as features [1]. This dual projection (BIPLOT) depicts the relation between objects and features and enables detection of clusters of correlated features as an indication of the presence of chemical components. Factor rotation [2] is therefore based on the observed cluster locations in the feature plot and yields the appropriate factor spectra. In addition, hierarchical clustering can be used in determining feature clusters.

This approach is demonstrated on a set of ammonia direct chemical ionization mass spectra obtained from the Algae spirulina cultivated in different locations in the world. Based on the BIPLOT a direction in the discriminant plot was found reflecting two homologous series of masses, one series being the unsaturated analogues of the other.

On account of the soft ionization method used, cluster ions rather than fragment ions can be expected to occur. Therefore the series of masses probably point to a number of homologous compounds which have a major impact on discrimination. High resolution measurements and MS-MS techniques will be used for identification of these compounds.

Calculations were carried out using the ARTHUR pattern recognition program [3] extended with routines for factor rotation [4] and a routine developed for display of features in factor and/or discriminant plots. Conversion of factor rotation data into the spectrum format of the SSX-software system [5] enables the use of this package for conventional mass spectral manipulation.

REFERENCES

1. K.R. Gabriel, Biometrika, 58 (1971) 453-467.

2. W. Windig, J. Haverkamp and P.G. Kistemaker, Anal. Chem., 55 (1983) 81-88.

3. Infometrix, Inc., Seattle, U.S.A.

4. Program for factor rotation kindly supplied by the FOM-Institute for Atomic and Molecular Physics, Amsterdam, The Netherlands.

5. Finnigan-MAT Bremen, F.R.G.

PATTERN RECOGNITION APPLIED TO DESIGN AND CONTROL OF CHEMICAL PROCESSING-MODEL SYSTEM STUDY

Wei-chuan Lai and Barbara Krieger
University of Washington
Department of Chemical Engineering
Seattle, Washington 98195

For organic solids, design and control of combusion or thermal conversion processes is complicated by the large number of low yield reaction products resulting from the initial degradation reaction. It is usually desired to narrow or optimize such a complex product distribution by careful choice or control of process variables. However, strategies for understanding and manipulating complex product slates are generally found with great difficulty during process development. Identification of an adequate but appropriately simple reaction model for process simulation purposes is equally demanding.

Multivariate analysis, particularly principal component analysis and discriminant analysis, has provided a rigorous basis for simplifying the product slate from a model chemical process without loss of information. The model process under study is the devolatilization of wood particles in gasifiers or stoker boilers. Process data include more than 50 time-dependent component concentrations (in the gas, liquid, and solid phases), time-temperature histories at several locations, and time-density histories within the particle, as well as gross reaction product yields. Preliminary results used to reduce the dimensionality of the data set are presented. Extensions of factor analysis techniques to provide appropriate lumping schemes for kinetic modeling purposes will also be described. Discriminant analysis, which can aid in reducing the number of process measurements that must be taken in order to precisely control the chemical reactor, will also be discussed.

DATA ANALYSIS OF CHEMICAL VAPORS USING SURFACE ACOUSTIC WAVE DEVICES AND PATTERN RECOGNITION METHODS

Susan Rose
Naval Research Laboratory
Chemical Division
Code 6110
Washington, D. C. 20375-5000

Surface acoustic wave devices with different vapor sensitive coatings are being used to detect low concentrations of vapors in air. Each sensor can be made to be highly sensitive, reversible, and reproducible. Individually, the sensors lack selectivity; however, an array of the sensors can produce a unique fingerprint for each vapor of interest. A data matrix, formed by exposing 12 coatings to 66 vapors representing different chemical classes and concentrations, has been studied using pattern recognition methods. Pattern recognition is one way of determining both the uniqueness of the information obtained by the array and the classification capacity of the sensors. Principal components analysis and clustering methods have been successful in investigating the clustering of the data produced by the sensors. Pattern recognition methods have also provided information useful in developing an understanding of the chemical interactions occurring between the coatings and the vapor species. In one application, supervised learning techniques were used to reduce to four the number of sensors necessary to separate 18 hazardous vapors from the others tested.

PALEO-ENVIRONMENTAL RECONSTRUCTION OF A TEXAS LIGNITE DEPOSIT BY NUMERICAL INTEGRATION OF PYROLYSIS MASS SPECTROMETRY, PYROLYSIS GAS CHROMATOGRAPHY AND CONVENTIONAL CORE CHARACTERIZATION DATA

Margriet Nip
Lab. Organische Geochemie
TH-Delft
de Vries van Heystplantsoen 2
Delft, The Netherlands

A set of extensively characterized drill core samples from a Texas Lignite Deposit (EOCENE) was analyzed by means of Curie-point pyrolysis mass spectrometry (PY-MS) (in combination with Curie-point pyrolysis gas chromatography mass spectrometry) and pyroprobe pyrolysis gas chromatography (Py-GC).

By means of factor, discriminant and canonical correlation analysis the pyrolysis data were numerically integrated with the conventional data of the samples. Because both the Py-MS and Py-GC data sets described the chemical differences between the samples in the most complete way, they were used as the basis for canonical correlation. As the Py-GC data set was not complete, the missing objects were replaced using a method by which the canonical variate for every variable in the Py-GC data set was calculated based on the scores of the overlapping part of the Py-MS data, subsequently the rotated scores for the complete set of discriminant scores were calculated making use of the linear combination resulting from canonical correlation.

The dimensionality of the canonical variate space (CV space) based on numerical integration of the Py-MS and Py-GC data sets was reduced to two dimensions (CV subspace) in which the major chemical tendencies in the original CV space were used as "Chemical Axes".

All variables of the conventional data sets were projected into the CV subspace by calculating the correlation coefficients of these variables with the standardized scores, resulting in the loadings of the mass variables.

Good correlations were observed between specific pyrolysis products and conventional data. Based on these correlations missing values in the conventional data sets of the samples were replaced. The correlations were used for the reconstruction of the paleoenvironment of deposition of the Texas Lignite Seam. This environment is very similar to the depositional environment which nowadays exists on the alluvial plains of the Mississippi Delta.

COMPUTER-ENHANCED INTERPRETATION OF CATALYTIC HYDROPROCESSING EFFECTS IN LOW VOLTAGE MASS SPECTRA OF COAL-DERIVED LIQUIDS

Koli Taghizadeh, Henk L.C. Meuzelaar
Biomaterials Profiling Center
University of Utah
391 S. Chipeta Way, Suite F
Research Park
Salt Lake City, Utah 84108

and

Burt Davis
Kentucky Center for Energy Research Laboratory
Iron Works Pike, P.O. Box 13015
Lexington, Kentucky 40512-3015

In order to assess the effect of catalyst aging and reactor temperature during the catalytic hydrotreatment of two coal-derived liquid in the Wilsonville pilot plant, low voltage mass specrometry (MS) and multivariate statistical analysis techniques were used.

It is nearly impossible to detect and monitor the minute changes in the composition of the coal liquids as a function of time and temperature without the use of multivariate statistical analysis techniques. Factor and discriminant analysis revealed a marked clustering of the coal liquid samples before as well as after hydrotreatment. The corresponding, numerically extracted, discriminant spectra illustrated that compound series such as octahydrophenanthrenes are relatively high in the hydrotreated product, whereas aromatic compound series such as biphenyls and or acenaphthenes are more prominent in the hydrotreater feed.

Canonical Correlation between run time and the corresponding mass spectra (can. corr. coeff. = 0.995) indicated that in the early days of the process more low molecular weight compounds (tetralins, phenols, naphthalenes) were formed, whereas with progress of time higher molecular weight (viz. polynuclear aromatic) materials became more dominant, possibly through condensation reactions.

Canonical correlation between complete data sets of low voltage mass spectra and "conventional" data (elemental analysis, ^{1}H NMR, solubility classes) produced on single canonical variate function (can. corr. coeff. = 0.995) which showed the effect of hydrotreatment. It was found that with progress of time, the hydrotreatment effect decreased gradually, probably due to catalyst aging. Attempts to obtain a clear indication of the effect of reactor temperature were unsuccessful due to spurious correlations between run time and reactor temperature as a result of poor experimental design.

MULTIVARIATE DATA ANALYSIS OF PYROLYSIS MASS SPECTRA

J. J. Boon, G. B. Eijkel, R. Hoggerbrugge and P. G. Kistemaker
FOM Institute for Atomic and Molecular Physics
Amsterdam, The Netherlands

Pyrolysis mass spectrometry (Py-MS) is generally applied to highly complex mixtures. This leads to correspondingly complex mass spectra. Direct interpretation of an individual spectrum is generally not possible.

In most applications interpretation is focused on the analysis of the differences in spectra of a series of related materials. The most descriptive differences are then more or less successfully correlated with varying concentrations of proposed chemical compounds in the sample material.

The retrieval of the significant differences and the correlation with Physico-chemical information is performed by multivariate data analysis techniques.

In this contribution the multivariate analysis approach as used in our laboratory will be presented. The procedures used are taken from the ARTHUR package. Additional procedures for discriminant analysis and canonical variate analysis were implemented as repeated principal component analysis. Some results of the multivariate analysis were compared with other analytical data on the samples as obtained by photo-ionization mass spectrometry, collision induced dissociation mass spectrometry and pyrolysis GC-MS. Three applications will be presented.

In the first project the depolymerization of straw cell walls is studied. The depolymerization of this poorly biodegradable material is induced by treatment with steam and anhydrous ammonia. In another project the FTE of plant polymers and their partially degraded fractions is studied in peat samples.

As a last example the analysis of mud from harbours and sea dumpsites is demonstrated.

Py-MS data and other sediment characteristics such as carbon content and heavy metal concentrations are correlated by canonical variate analysis.

SPECTROSCOPIC ANALYSIS OF PYROLYSIS TARS FROM FOUR WESTERN COALS OF DIFFERENT RANK USING AN INTEGRATED DATA REDUCTION APPROACH

Barbara L. Hoesterey, Henk L.C. Meuzelaar

The compositional complexity of coal liquids due, of course, to the structural complexity of coal itself, challenges the analyst to provide a complete yet concise description of physio-chemical properties useful for modeling and predictive value. Coal liquids (tars) are generally viscous liquids, soluble in carefully chosen solvents, and are thereby amenable to analysis by various chemical analytical procdures.

Tars produced by a Wellman Galusha fixed bed gasifier from four Western coals, viz., Hiawatha #2 (hvBb), Upper Hiawatha (hvCb), Adaville (subbituminous) and Beulah Gap (lignite) were analyzed by Low Voltage Mass Spectrometry (MS), Fourier Transform Infrared Spectroscopy (IR), Proton Nuclear Magnetic Resonance Spectrometry (NMR) and Conventional techniques (CONV) including elemental analysis and liquid chromatography. Factor (or discriminant) analysis was performed on each data set (MS, IR, NMR and CONV) separately for data reduction. The data sets were then subjected to canonical correlation analysis to find the common information from all techniques. Evaluation of the integrated MS, IR and NMR data showed Hiawatha tar to be highest in aliphatic and aromatic hydrocarbons and Beulah Gap tar to contain large amounts of phenolic and dihydroxybenzene moieties. Canonical correlation of conventional data from the tars with other data sets showed hydrogen content and total tar yield (tar coal) to correlate srongly with rank.

Pyrolysis MS data and conventional data on the feedstock coals were also correlated with the tar data. In fact, in previous work a high degree of correspondence was reported between the composition of a caol pyrolysis tar produced in a 1.5 ton/day fixed bed gasifier and the pyrolysis pattern of a 20 g sample of the same coal heated directly in front of the ion source of a mass spectrometer.

From the results obtained using an integrated data analysis approach, it is possible to begin to construct models to predict tar yield and composition from knowledge of feedstock coal properties, resulting in tremendous time and cost savings. In addition, greater understanding of pyrolysis reaction mechanisms can help to optimize process design and operating procedures.

APPLICATIONS OF A SPECTRAL WORKSTATION IN INFRARED SPECTROSCOPY

S.R. Lowry and G.L. Ritter
Nicolet Instrument Corporation
5225 Verona Road
P.O. Box 4508
Madison, Wisconsin 53711-0508

Now that most high-performance infrared spectrometers are computerized and producing high-quality digital data, a large number of spectral analysis methods can be applied. Nicolet Instrument Corporation has developed a spectral workstation with optimized interactive display processor that is specifically designed for infrared spectral analysis.

We will be demonstrating a number of application software packages that have been implemented on the workstation. Some of these packages are described below:

<u>Infrared Spectral Search System</u>: This uses a full spectral fit algorithm to determine the most similar spectra in a reference library. The system uses a CD ROM disk to store the actual reference spectra.

<u>Spectral Deconvolution</u>: This program uses inverse transform techniques to deconvolve line shape information from a spectrum and enhance the "resolution" of the spectrum.

<u>Factor Analysis</u>: A simple factor analysis program is available to determine the number of degrees of variance in a set of spectrum.

<u>Spectral Interpretation Program</u>: This software is based on the work of Woodfuff, et. al., and provides an "expert system" approach to infrared spectral interpretation.

<u>PLS Quantitative Analysis</u>: This is an infrared spectral quantitative analysis package based on the partial least squares algorithm. The software is completely modular with a "spread sheet" data entry system and report generator.

<u>Stucture Generation Program</u>: We have implemented a structure generating program as part of our long-range data base project. This will permit structures to be retrieved and displayed as part of the specral search system.

<u>Curve Analysis Package</u>: This is a program that allows the user to generate a number of peaks and to then fit them to an experimental spectrum. This is an interactive program with full overlay display capabilities.

<u>Basic Spectral Analysis</u>: This is the program that performs the basic spectral manipulation functions such as: subtraction, baseline correct, derivative, peak pick, and smoothing.

USING "MATHEMATICAL CHROMATOGRAPHY" TO FIND PURE MASS SPECTRA IN DOPING ANALYSIS

Erkki J. Karjalainen
United Laboratories Ltd.
P. O. Box 70
00511 Helsinki, Findland

The use of drugs to improve physical performance in sports is forbidden by the rules of the International Olympic Committee (IOC). United Laboratories has done analyses for doping control since 1977. It was the first laboratory in Scandinavia to receive the accreditation by IAAF and IOC to perform these tests. Most of the test methods are based on GC and GC/MS. I have developed a set of computer programs in FORTRAN IV for processing full mass spectra collected by continuous scanning in GC/MS. The term "mathematical chromatography" attempts to convey the central idea of the method: If the chromatographic resolution is not sufficient, "separation" is mathematically achieved by computer processing of the mass spectra. The sum spectra contain more than enough information to find the pure components and their concentration profiles.

The algorithm starts by assuming random spectra and solves for the concentrations. Then again, the spectra are solved on the basis of the concentrations. The iteration is repeated 10 to 20 times to converge to a stable solution. The number of components in a given chromatographic run is found empirically. When the number of fitted components gets too large, the iterations diverge instead of converging. Components having (nearly) identical concentration profiles still remain unseparated. A given component should be present in at least four spectra in the observations. Each mass number must have at least one observation with an intensity of zero. The computer used is an Eclipse S/250 with the array processor FPS-100, which performs floating point arithmetic at 10 megaflops.

The results have been useful in doping analysis during the last five years. The steroid fraction used in the analysis for anabolic steroids has about three times as many components as visible peaks in the total ionization. The advantages of "mathematical chromatography" can be summarized as follows:

- The method is objective, no previous knowledge about spectra is needed.
- No previous knowledge about concentrations or peak shapes is needed.
- The process produces a "mass balance" of all ions in the GC/MS run.
- All components exceeding background noise are found.
- The analysis by computer takes as long as the GC/MS run itself.
- The method is easily modified for other 2-D spectroscopies.

References:

Karjalainen, E. J., Karjalainen, U. P. "Mathematical Chromatography" - Resolution of Overlapping Spectra in GC-MS. Medical Informatics Europe 85, Proceedings, eds. Roger FH, Gronroos P, Tervo-Pellikka R, O'Moore R, Springer Verlag, 1985; 572-8.

AN MICRO-PROCESSOR CONTROLEED INTELLIGENT DIGITIZER FOR SPECTROGRAMS

Cheng Qian, Yizhao Wang, Xianghong Yan
Shufeng Zhu and Yi Song
Shanghai Institute of Organic Chemistry
345 Ling Ling Road
Shanghai, China

At the Shanghai Institute of Organic Chemistry, a computerized IR data bank has been available to chemists since 1985, which now incorporates 35,000 spectra and more than 20,000 associated structures. The information for an IR spectrum includes wavelength, intensity and half-height width of absorption peaks. However, the manual collection of spectral data is a very tedious work which causes relatively high error rate and almost impossible to be further applied to full spectrum digitization. An intelligent digitizer, consisting of a drum scanner and a Z-80 micro-processor has been made. The micro-processor collected the scanner's output data, recognized the spectrum profile and separated it from the grid background so that the "clean" spectral data can be achieved automatically.

The data processing algorithm is based on filtering and heuristic search which effectively eliminate the vertical and horizontal background grids. A suitable data structure led to very compact storage of image data from a spectrogram, characteristic of 2 gray-levels and low black/white ratio, due to its composition of narrow lines and curves. The discriminant function of the heuristic search took into account the length of a single vertical black segment, the total length of all black segments on a vertical scanning line, the neighbouring situation of a segment with regards to each segment of the previous vertical scanning line etc.

The digitizer can be connected to any type of mainframe, mini or micro computer as long as it has a RS-232C port and its operation can be controlled by the host using ASCII coded commands. It can be utilized for digitization of IR spectra and other histogram-like graphs.

A SOFTWARE PACKAGE FOR MULTIVARIATE DATA DISPLAY AND CLUSTER ANALYSIS ON THE IBM PC.

Stephen, L. Morgan, Michael D. Walla and Michael Abdalla
Department of Chemistry
University of South Carolina
Columbia, SC 29208

Computer-assisted processing is essential for the rapid analysis and interpretation of complex data generated by a variety of chromatographic and spectroscopic analytical methods. This poster presents an integrated software package designed for the IBM Personal Computer for the pre-treatment and display of multivariate data using a variety of pattern recognition techniques including hierarchical cluster analysis, nonlinear mapping, and principal component analysis. The package also includes subroutines for calculating statistics descriptive of multivariate data to aid in feature selection, as well as routines implementing a variety of transformations such as normalizing and/or autoscaling. The programs are written in Standard FORTRAN and use a minimum of machine specifix graphics capabilities so as to retain portability to other computer environments. Examples of data treatment, display, and cluster analysis will include capillary gas chromatography and pyrolysis GC-MS applications.

A QUANTITATIVE MEASURE OF LIBRARY SEARCH RELIABILITY

T.L. Isenhour and P.B. Harrington
Utah State Universi.y
Logan, UT 84321

Spectral libraries are important for structure elucidation of complex molecules. Mass spectroscopy (MS) and infrared spectroscopy (IR) are common techniques used for molecular structure determination. The utility of library searching may be reflected in the sizes to which these databases have grown, 79,560 entries for MS and over 95,000 entries for IR (1,2). Search results of large libraries will produce lists of spectrally nearly identical compounds. A measure of search performance will facilitate the interpretation of the search results.

Intra-library search results contain information on how the library search performs under ideal circumstances. This information in the past has been used to evaluate library configurations (3,4). This same information may also be used for determining the quality of a spectral match between a target spectrum containing noise and the reference spectrum in the library. A non-probalistic quantitative measure of the reliability of spectral matches has been developed. This metric may perform better than the current methods for evaluating spectral libraries.

CITATIONS

1. McLafferty, F.W., Stauffer, D.B.; Int. J. Mass Spectrom. Ion Phys. 1984, 58, 139.
2. Sadler research Libraries, Spectroscopy. 1986, 1, 4.
3. Rasmussen, G.T., Isenhour, T.L. J. Chem. Inf. Comput. Sci. 1979, 43, 1362.
4. Delaney, M.F., Warren, F.V., Hallowell, J.R., Jr. Anal. Chem. 1983, 55, 1925.

ABSOLUTE QUALITATIVE AND QUANTITATIVE ANALYSIS OF PYROLYSIS MASS SPECTRA OF MIXTURES

Willem Windig, William H. McClennen and Henk L.C. Meuzelaar
Biomaterials Profiling Center
391 S. Chipeta Way, Suite F, Research Park
Salt Lake City, Utah 84108

The paper describes factor analysis of the "correlation around the origin" matrix as applied to (pyrolysis) mass spectrometry data. This approach makes it possible to calculate the spectra of pure components from a data set of mixtures in which these pure components are not present. Furthermore, the absolute concentration of the components in the mixtures can be calculated. Examples will be given of results obtained on data sets consisting of (pyrolysis) mass spectra from biopolymers, jet fuels and technical polymers.

Participants

LIST OF PARTICIPANTS

Lawrence J. Altman
Mobil R&D Corp.
Billingsport Road
Paulsboro, NJ 08066

Ernest Armstrong
The Standard Oil Co.
4440 Warrensville Center Road
Cleveland, Ohio 44128

Hal Brand
Lawrence Livermore
P.O. Box 808 L-308
Livermore, CA 94550

Ernest H. Braue, Jr.
USAMRICD
Analytical Chemistry Branch
BLDG E-3100
Aberdeen Proving Ground, MD 21010

James M. Brown
Exxon Research & Eng. Co.
Clinton Township, Route 22 East
Annandale, NJ 08801

James W. Bunger
Research Asst. Professor
Fuels Engineering
320 Browning Bldg.
Univ. of Utah
Salt Lake City, UT 84112

Tamnoy Chakravarty
Univ. of Utah
Biomaterials Profiling Center
391 S. Chipeta Way Suite F
Salt Lake City, UT 84108

Jean T. Clerc
University of Berne
Pharmaceutical Institute
Baltzerstrasse 5
CH-3012 Berne
SWITZERLAND

Michael J. Coggiola
SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025

Anders Colmsjo
University of Stockholm
Dept. Analytical Chemistry
S-10691 Stockholm

Jack Comeford
CHEM LAB-US Army DPG
2549 E. Wilmington Ave.
Salt Lake City, UT 84109

Olof Dahlman
Swedish Tobacco Co.
Box 17007
Stockholm, S-10462
SWEDEN

F. Delaroff
Roussel-UCLAF
102 Route de Noisy
93230 Tomainville
FRANCE

Fred L. DeRoos
Battelle
505 King Ave.
Columbus, Ohio 43201

Doug Dorman
Lilly Research Labs
Dept. Mc 525
Indianapolis, Ind. 46285

Steve L. Durfee
Colo. School of Mines
Dept. of Chemistry
Golden, CO 80401

Nancy Ryan Gray
Exxon Production Research
P.O. Box 2189 South 187
Houston, TX 77252

Peter R. Griffiths
Department of Chemistry
University of California
Riverside, CA 92521

John Halliday
Analytical Branch
Physical Sciences Division
Dept. of the Army
Natick Research Devl. & Eng. Cnt.
Natick, MA 01760-5020

Peter B. Harrington
Utah State University
Dept. of Chem. & Biochem.
Logan, UT 84322

B. Austin Haws
Utah State University
Biology Dept.
Logan, UT 84322-5305

Stephen R. Heller
USVA-ARS, ASRI-MDCL
Room 233-C
Bldg. 005 BARC-West
Beltsville, MD 20705

Barbara Hoesterey
Univ. of Utah
Biomaterials Profiling Center
391 S. Chipeta Way Suite F
Salt Lake City, UT 84108

Peter W. Hovey
University of Dayton
300 College Park
Dayton, OH 45469

Fu Yu Hshieh
University of Utah
Fuels Engineering Dept.
Salt Lake City, UT 84108

Thomas L. Isenhour
Utah State University
Dept. of Chem. & Biochem.
Logan, UT 84322-0300

Jumpol Jaturapitpornsakul
University of Utah
Dept. of Fuels Eng.
Salt Lake City, UT 84112

Kiyokatsu Jinno
Toyohashi University of
Technoloty
Toyohashi 440
JAPAN

Peter C. Jurs
Penn State University
152 Davey Lab-Chemistry Dept.
University Park, PA 16802

Robert Kajioka
Public Health Labs
Box 9000 Terminal A
Toronto, Ontario
CANADA

James W. Kauffman
Honeywell Inc.
MS 665-M
13350 U.S. Hwy. 19 S
Clearwater, FL 33546

E.R. Karjalainen
United Laboratory
P.O. Box 70
00511 Helsinki
FINLAND

Marwin K. Kemp
Amoco Production
P.O. Box 3385
Tulsa, OK 74102

Richard M. King
Waters Div. Millipore
Maple St.
Milford, MA 01757

Piet J. Kistemaker
FOM Institute for Atomic &
Molecular Physics
Kruislaan 407 Amsterdam Wgm
THE NETHERLANDS

Anthony J. Klimasara
GTE Products Corp.
100 Endicott Street
Danvers, MA 01923

C.B. Koons
Exxon Production Res. Co.
P.O. Box 2189
Houston, TX 77001

Barbara B. Krieger
Benson Hall BF-10 Dept. Chem Eng.
University of Washington
Seattle, Washington 98195

Ned Lagin
Bio Rad Laboratories
1414 Harbour Way South
Richmond, CA 94804

Stephen R. Lowry
Nicolet INst.
5225-1 Verona Road
Madison, WI 53711

George Lozos
Allied Bendix ESD
1400 Taylor Ave.
Baltimore, MD 21284

Hal MacFie AFRC Institute of
Food Research
Bristol Laboratory
Langford, Bristol
ENGLAND

Edmund Malinowski
(E.R.M. - Stevens Inst. of Tech.)
49 New England Dr.
Lake Hiawatha, NJ 07034

John C. Marshall
Utah State University
Dept. of Chemistry & Biochem.
Logan, UT 84322-0300

Charles Mayne
University of Utah
Post Doctoral Rsc. Assoc.
B23 Hnry Eyrng Bld.
Salt Lake City, UT 84112

Sue Mazer
Univ. of Dayton Research Inst.
300 College Park
Dayton, OH 45469

William H. McClennen
Univ. of Utah
Biomaterials Profiling Center
391 S. Chipeta Way Suite F
Salt Lake City, UT 84108

Henk L.C. Meuzelaar
Univ. of Utah
Biomaterials Profiling Center
391 S. Chipeta Way Suite F
Salt Lake City, UT 84108

Stephen L. Morgan
Department of Chemistry
University of South Carolina
Columbia, SC 29208

Tormod Naes
Norwegian Food Research
Box 50
1432 AS-NLH
NORWAY

Sharon L. Neal
Emory University
Chemistry Dept.
Atlanta, GA 30021

Margriet Nip
Delft Univ. of Technology
Department of Chemistry and
Chemical Engineering; Organic
Geochemistry Unit
De Vries van Heystplantsoen 2
2628 RZ Delft
THE NETHERLANDS

Marc R. Nyden
U.S. Dept. of Commerce
National Bureau of Standards
Gaithersburg, MD 20899

Paul Painter
Penn State University
325 Steidle
University Park, PA 16802

Nantelle S. Pantaleo
Amoco Production
P.O. Box 3385
Tulsa, OK 74102

Joseph A. Platt
John Hopkins University
Allied Physics Lab.
Laurel, MD 20707

Ronald J. Pugmire
Assoc. Vp. Research
304 Park Bldg.
University of Utah
Salt Lake City, UT 84112

Cheng Qian
Shanghai Institute of Organic Chem.
345 Ling Ling Road
Shanghai,
CHINA

Joseph M. Richards
Univ. of Utah
Biomaterials Profiling Center
391 S. Chipeta Way Suite F
Salt Lake City, UT 84108

Garry L. Ritter
Nicolet Instrument Co.
5225-1 Verona Road
Madison, WI 53711-0508

Marc E. Rippen
Pratt and Whitney
P.O. Box 2691
WPB, Fla 33402

Joseph Roehl
ALLIED/Bendix
1400 Taylor Ave.
P.O. Box 9840
Baltimore, MD 21284-9840

Susan Rose
Naval Research Lab.
Chem. Div. Code 6110
Washington, D.C. 20375-5000

Abraham Savitzky
Silvermine Resources Inc.
3 Mail Coach Court
Wilton, Connecticut 06897

Donald R. Scott
Environmental Protection Agency
ERC Annex Building, MD-78A
Research Triangle Park, N.C. 27711

A. Peter Snyder
Chem. Research, Dev. & Eng. Center
AMCCOM, SMCCR-RSL
Aberdeen Proving Ground, MD 21010-5423

Grant Gill Smith
Utah State University
Chemical & Biochemical
Logan, UT 84322-0300

Michael Spragg
Perkin-Elmer Limited
Post Office Lane, Beaconsfield
Buckinghamshire, HP9 1QA
ENGLAND

D.G. Streefkerk
Royal Dutch/Shell Laboratory
3003, 1003 AA Amsterdam
THE NETHERLANDS

Henk Stolk
Univ. of Utah
Biomaterials Profiling Center
391 S. Chipeta Way Suite F
Salt Lake City, UT 84112

Dennis Suiter
IBM Inst.
611 Anton Blvd; 8th floor
Costa Mesa, CA 92626

Koli Taghizadeh
Univ. of Utah
Biomaterials Profiling Center
391 S. Chipeta Way Suite F
Salt Lake City, UT 84108

Albert C. Tas
TNO-CIVO Food Analysis Institute
Utrechtseweg 48/P.O. Box 360
3700 AJ Zeist
THE NETHERLANDS

Charles L. Wilkins
Chairman, Department of Chemistry
University of California at Riverside
Riverside, CA 92521

Willem Windig
Univ. of Utah
Biomaterials Profiling Center
391 S. Chipeta Way Suite F
Salt Lake City, UT 84108

Carla M. Wong
Lawrence Livermore Natl. Lab.
P.O. Box 808 L-365
Livermore, CA 94550

Hugh B. Woodruff
Merc Sharp & Dohme Research
Laboratory
P.O. Box 2000
Rahway, NJ 07065

S. Yegnasubramanian
AT&T Bell Labs.
Room 1A-373
600 Mountain Ave.
Murray Hill, NJ 07974

Richard A. Yost
Dept. of Chemistry
Univ. of Florida
Gainesville, FL 32611

EXHIBITORS

Biomaterials Profiling Center; University of Utah

The Biomaterials Profiling Center is a widely recognized leader in the area of chemical characterization of complex organic materials as well as in the field of computerized analysis of measurement data. Present software development efforts center around a System for Interactive Graphics-oriented Multivariate Analysis (SIGMA). SIGMA is a user friendly program developed for interactive graphics oriented multivariate analysis. As such it has extensive graphics and multivariate analysis capabilities.

IBM; IBM Instruments

IBM Instruments has integrated advanced optics, innovative software, versatile accessories and the powerful IBM Personal Computer AT to deliver exceptional precision, productivity and reliability in benchtop spectroscopy. The accuracy and precision of the IR/44 give you maximum confidence in your IR results. The unique optical design provides excellent system stability for repeatable analysis - scan to scan, run to run, day to day. Our advanced chemometrics allow you to perform complex spectral arithmetic automatically, assuring you of consistent and accurate results. The IR/44 attains a high level of productivity through fast analyses, easy-to-use instrumentation and comprehensive automation features. Highspeed focused optics maximize optical throughput for fast analysis, even of your most demanding low energy samples. Innovative software design makes the IR/44 easy to learn and convenient to use for both the occasional and experienced spectroscopist. And the IR/44 offers comprehensive automation capabilities for unattended measurements, quantification and reporting.

Nicolet Instrument Corp.

Spectral workstation with full Infrared Spectral Software Package. The workstation functions include; spectral manipulation, quantitative analysis, LIMS, spectral interpretation and curve analysis.

END

11-86

DTIC